

Entropy for theories with indefinite causal structure

Sonia Markes*

*Department of Applied Mathematics, University of Waterloo, Waterloo, Ontario, Canada, N2L 3G1 and
Perimeter Institute for Theoretical Physics, Waterloo, Ontario, Canada, N2L 2Y5*

Lucien Hardy†

Perimeter Institute for Theoretical Physics, Waterloo, Ontario, Canada, N2L 2Y5

Entropy is a concept that has traditionally been reliant on a definite notion of causality. However, without a definite notion of causality, the concept of entropy is not all lost. Indefinite causal structure results from combining probabilistic predictions and dynamical space-time. Combining the probabilistic nature of quantum theory and dynamical treatment space-time from general relativity is an approach to the problem of quantum gravity. The causaloid framework lays the mathematical groundwork to be able to treat indefinite causal structure. In this paper, we build on the causaloid mathematics and define a causally-unbiased entropy for an indefinite causal structure. In defining a causally-unbiased entropy, there comes about an emergent idea of causality in the form of a measure of causal connectedness, termed the Q factor.

I. INTRODUCTION

In Newtonian physics, physical processes are understood with respect to a fixed spatial coordinate system and a time parameter, which is absolute and ever increasing. Predictions are entirely deterministic. Quantum theory and general relativity depart from this classical picture in opposing manners. Quantum theory gives probabilistic predictions as to the outcomes of measurements, but retains fixed space and time coordinates. On the other hand, general relativity is deterministic, but shows that space and time form a dynamical structure. Reconciling these fundamental philosophical differences is one of the many challenges one is faced with in trying to construct a theory of quantum gravity. There have been many different approaches to this problem with many different results [1, 2, 3, 4, 5, 6]. One way of moving forward is to dismiss classical assumptions and create a probabilistic theory that has a dynamic causal structure. However, what results is indefinite casual structure. This is more radical than either probabilistic predictions or dynamical space-time structure. In general relativity, a separation between space-time locations is either space-like or time-like. An indefinite causal structure would allow for a separation between space-time locations to be something like a quantum superposition of a space-like and a time-like separation. While we may be uncertain of the causal structure of the path between measurements, we know where in space-time we make measurements, what measurements we have made, and what outcomes we get. With this data, we can examine probabilistic correlations for information. The causaloid framework ([7],[8],[9]) provides us with the necessary structure. We will outline the essentials of this framework in Section 2.

It is natural in discussions of causal structure to raise the question of entropy. The second law of thermodynamics tells us that in an isolated system, entropy can increase or remain the same, but it can never decrease [10]. In information theory, entropy is viewed as being a measure of uncertainty before we measure a state or equivalently, the amount of information gained by upon learning the state of a system [11]. Inherent in both concepts of entropy is an assumed causal structure, specifically that there exists a background time. The standard definition of entropy is in the context of a definite causal structure with reference to absolute time. In order to make sense of entropy in an indefinite causal structure, a clear definition must be established. To do so requires consideration of the following questions:

What are the concepts from the usual picture of entropy in a definite causal structure that are necessary to define entropy? What are the analogues to these concepts in a picture with indefinite causal structure?

Using the formalism introduced in the causaloid framework, we are able to provide answers to these questions and then, define a causally-unbiased entropy.

*Electronic address: smarkes@perimeterinstitute.ca

†Electronic address: lhardy@perimeterinstitute.ca

In Section 2, we will review the relevant aspects of the causaloid framework. We then proceed with the new developments. In Section 3, we define a new type of product that is utilized in the work on entropy. The definition of causally-unbiased entropy and resulting features are developed in Section 4.

II. CAUSALOID FRAMEWORK

A. The Picture

Every experiment results in a set of data from making measurements on a system. Each piece of data could be thought of as a card with three pieces of information on it; where the measurement is made in space-time, what is measured, and what the result of the measurement is. We will represent each card (or piece of data) as (x, f_x, y_x) where x denotes the space-time information, f_x denotes the information pertinent to a choice of measurement or action, and y_x denotes the information regarding an observation or outcome of a measurement. The set of all possible cards (i.e. all possible measurements with all possible outcomes with every space-time configuration) is denoted V . We can imagine running an experiment an infinite number of times so as to be able to obtain relative frequencies. In order for the cards to tell us the relative frequencies, we must systematically sort them.

Each distinct x is defined as an *elementary region* of space-time. A *composite region*, denoted \mathcal{O}_1 , is a set of elementary regions. (Note: These definitions of “elementary region” and “composite region” differ from those in [7, 8, 9].) Therefore, these cards can be sorted according to their associated space-time region. The set of all possible cards with the same space-time information x written on them is the *measurement information for elementary region* x . We denote this set as R_x . The measurement information for composite region \mathcal{O}_1 is the union of all sets of measurement information for the elementary regions contained within the composite region. More concisely,

$$R_1 \equiv \bigcup_{x \in \mathcal{O}_1} R_x \quad (1)$$

We can further sort the measurement information in a region. The *procedure in a region*, denoted F_x , is the set of all distinct choices of measurement recorded for the region x .

$$F_x \equiv \bigcup_{all\ f_x, y_x} (x, f_x, y_x) \quad (2)$$

For composite regions, we define the following set:

$$F_1 \equiv \bigcup_{x \in \mathcal{O}_1} F_x \quad (3)$$

Similarly, the *outcome set in a region*, denoted Y_x is defined to be the set of all distinct outcomes of a measurement recorded for the region x .

$$Y_x \equiv \bigcup_{all\ y_x} (x, f_x, y_x) \quad (4)$$

Again, for composite regions, we define

$$Y_1 \equiv \bigcup_{x \in \mathcal{O}_1} Y_x \quad (5)$$

Notice that $Y_x \subseteq F_x \subseteq R_x$ and $Y_1 \subseteq F_1 \subseteq R_1$. The composite structure we expected of our space-time regions is reflected structure throughout these sets. A set of cards with the measurement information for a region has no more or less structure than an elementary region of space-time. Therefore, without adding structure or losing generality, we can take the sets R_x to be elementary regions, at least, for the purposes of this paper. From this point forward, the term *region* will be used interchangeably to refer to objects of type x or \mathcal{O}_1 and type R_x or R_1 .

Notice that the set of all cards V can be viewed as all the cards from all (elementary) regions.

$$V = \bigcup_{all\ x} R_x \quad (6)$$

So V is the largest of all regions that can be considered.

These definitions provide a firm foundation on which the causaloid framework rests both mathematically and conceptually.

B. First level physical compression

The most basic quantity that we would want to be able to calculate is the probability that a certain (set of) outcome(s) is observed given that a certain (set of) measurement(s) has been performed at a (set of) location(s) in space and time. Suppose that the set of locations we are interested in is \mathcal{O}_1 . The set of all the cards corresponding to these locations called R_1 . We write pairings of measurements and corresponding outcomes in R_1 as (Y_1, F_1) . A specific outcome and measurement pair is denoted as α_1 (or equivalently, $(Y_1^{\alpha_1}, F_1^{\alpha_1})$). The set of all α_1 in region \mathcal{O}_1 is Υ_1 . The set comprised of all the cards not in R_1 is $V - R_1$. We call (Y_{V-R_1}, F_{V-R_1}) the *generalized preparation* because it is the information that surrounds R_1 not only from the immediate past, but from the future and the rest of space-time as well. By the choices we make in setting up the experiment, we can put conditions on the generalized preparation such that $\text{Prob}(Y_V|F_V)$ is well-defined. (See Ref.[7] for details.) Then we can write

$$\text{Prob}(Y_V|F_V) = \text{Prob}(Y_1, Y_{V-R_1}|F_1, F_{V-R_1}) \quad (7)$$

For a specific pair $\alpha_1 \Leftrightarrow (Y_1^{\alpha_1}, F_1^{\alpha_1})$, we can write this probability as

$$\text{Prob}(Y_1^{\alpha_1}, Y_{V-R_1}|F_1^{\alpha_1}, F_{V-R_1}) \quad (8)$$

We use the short-hand p_{α_1} to denote the probability defined in Eq.(8). One way to specify the state of a system is to list all the possible p_{α_1} for elements of R_1 .

$$\begin{pmatrix} \vdots \\ p_{\alpha_1} \\ \vdots \end{pmatrix} \quad \alpha_1 \in \Upsilon_1 \quad (9)$$

However, this over-specifies the state. We do not usually need to know the probability of every outcome of every measurement in order to determine what the complete state of the system is. Physical theories tell us what relationships exist between variables and what constraints those relationships place on the variables of the system. These relationships and constraints can be used to determine a reduced set of probabilities from which all other probabilities can be represented. The reduced set of probabilities is defined such that any probability can be written as a linear combination of the probabilities in the reduced set. Let us denote the reduced or fiducial set in R_1 as $\Omega_1 \subseteq \Upsilon_1$. This process of going from the set of all the probabilities to the smallest essential set we call *first level physical compression*. This can be expressed as

$$\mathbf{p} = \begin{pmatrix} \vdots \\ p_{l_1} \\ \vdots \end{pmatrix} \quad l_1 \in \Omega_1 \subseteq \Upsilon_1 \quad (10)$$

such that

$$p_{\alpha_1} = \mathbf{r}_{\alpha_1} \cdot \mathbf{p} \quad (11)$$

where \mathbf{r}_{α_1} encodes the physical compression and therefore, is determined by the details of the physical theory. We can define a *decompression matrix*, $\Lambda_{\alpha_1}^{l_1}$ such that

$$\Lambda_{\alpha_1}^{l_1} \equiv \mathbf{r}_{\alpha_1}|_{l_1} \quad (12)$$

where $\mathbf{r}_{\alpha_1}|_{l_1}$ means the l_1 component of \mathbf{r}_{α_1} .

C. Second level physical compression and the causaloid product

Let us consider two distinct regions $R_1, R_2 \subset V$. In a similar fashion to the single region case,

$$p_{\alpha_1\alpha_2} = \text{Prob}(Y_1^{\alpha_1}, Y_2^{\alpha_2}, Y_{V-R_1-R_2}|F_1^{\alpha_1}, F_2^{\alpha_2}, F_{V-R_1-R_2}) \quad (13)$$

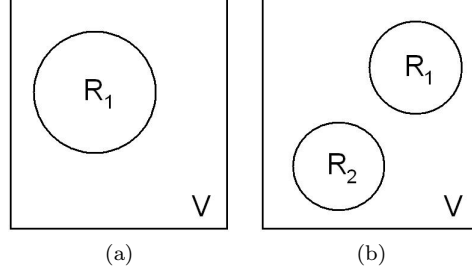


FIG. 1: (a) A single region R_1 in V (b) Two regions R_1 and R_2 in V

We specify the state of the system by listing all $p_{\alpha_1\alpha_2}$.

$$\begin{pmatrix} \vdots \\ p_{\alpha_1\alpha_2} \\ \vdots \end{pmatrix} \quad \alpha_1\alpha_2 \in \Upsilon_1 \times \Upsilon_2 \quad (14)$$

where \times is the cartesian product. It can be shown that

$$p_{\alpha_1\alpha_2} = \sum_{l_1l_2 \in \Omega_1 \times \Omega_2} \Lambda_{\alpha_1}^{l_1} \Lambda_{\alpha_2}^{l_2} p_{l_1l_2} \quad (15)$$

which implies that the following list of probabilities is sufficient.

$$\begin{pmatrix} \vdots \\ p_{l_1l_2} \\ \vdots \end{pmatrix} \quad l_1l_2 \in \Omega_1 \times \Omega_2 \subseteq \Upsilon_1 \times \Upsilon_2 \quad (16)$$

This is effectively first level compression on each index. But if a physical theory has some connection between the two regions, $\Omega_1 \times \Omega_2$ may no longer be the smallest set that is sufficient to represent all possible states. Then *second level physical compression* is possible. It is defined to be

$$\mathbf{p} = \begin{pmatrix} \vdots \\ p_{k_1k_2} \\ \vdots \end{pmatrix} \quad k_1k_2 \in \Omega_{12} \subseteq \Omega_1 \times \Omega_2 \quad (17)$$

such that

$$p_{\alpha_1\alpha_2} = \mathbf{r}_{\alpha_1\alpha_2} \cdot \mathbf{p} = \sum_{k_1k_2 \in \Omega_{12}} \mathbf{r}_{\alpha_1\alpha_2} \big|_{k_1k_2} p_{k_1k_2} \quad (18)$$

When $\Omega_{12} = \Omega_1 \times \Omega_2$, second level compression is trivial. But it is proven in [8] that it is possible that $\Omega_{12} \subset \Omega_1 \times \Omega_2$.

Now we can define a second level decomposition matrix. By comparing Eq.(15) and Eq.(18), we infer that

$$\mathbf{r}_{\alpha_1\alpha_2} \big|_{k_1k_2} = \sum_{l_1l_2 \in \Omega_1 \times \Omega_2} \Lambda_{\alpha_1}^{l_1} \Lambda_{\alpha_2}^{l_2} \Lambda_{l_1l_2}^{k_1k_2} \quad (19)$$

where

$$\Lambda_{l_1l_2}^{k_1k_2} = \mathbf{r}_{l_1l_2} \big|_{k_1k_2} \quad (20)$$

which is the desired second level decomposition matrix. This matrix encodes how we move from $p_{l_1l_2}$'s to $p_{k_1k_2}$'s. Using the definition of the first level decomposition matrix, Eq.(19) becomes

$$\mathbf{r}_{\alpha_1\alpha_2} \big|_{k_1k_2} = \sum_{l_1l_2 \in \Omega_1 \times \Omega_2} \Lambda_{l_1l_2}^{k_1k_2} \mathbf{r}_{\alpha_1} \big|_{l_1} \mathbf{r}_{\alpha_2} \big|_{l_2} \quad (21)$$

This defines the *causaloid product*, denoted $\mathbf{r}_{\alpha_1} \otimes^\Lambda \mathbf{r}_{\alpha_2}$ which unifies the different causal structure-specific products. Explicitly,

$$\mathbf{r}_{\alpha_1} \otimes^\Lambda \mathbf{r}_{\alpha_2} = \mathbf{r}_{\alpha_1 \alpha_2} \quad (22)$$

It is this product that allows us to look at the probabilistic correlations between arbitrary locations in space-time without specifying the causal relationship.

We have shown second level compression for the case where we have two regions. This is easily generalized for any number of regions. The object that would encode the compression for three regions would be $\Lambda_{l_1 l_2 l_3}^{k_1 k_2 k_3}$, for four regions would be $\Lambda_{l_1 l_2 l_3 l_4}^{k_1 k_2 k_3 k_4}$, etc. After second level compression over multiple regions, we have

$$\begin{pmatrix} \Lambda_{\alpha_1}^{k_1} \\ \Lambda_{l_1 l_2}^{k_1 k_2} \\ \Lambda_{l_1 l_2 l_3}^{k_1 k_2 k_3} \\ \vdots \end{pmatrix} \quad (23)$$

There is a third level of physical compression that compresses these multi-region Λ -matrices to give the *Causaloid*, Λ , which is defined as

$$\Lambda \equiv \left(\{\Lambda\} \mid \{\Lambda\} \subseteq \{\Lambda_{\alpha_1}^{k_1}, \Lambda_{l_1 l_2}^{k_1 k_2}, \dots\} \right) \quad (24)$$

where $\{\Lambda\}$ is determined by the rules of the physical theory (for detailed discussion of how this works see [2]). By decompressing the set $\{\Lambda\}$, we can obtain the Λ -matrix for any set of regions. This means that the Causaloid gives us the ability to perform any calculation that the physical theory allows for.

D. Well-defined probabilities

Up to this point we have exclusively dealt with probabilities conditioned on procedures. It is more useful to also be able to condition on outcomes. Specifically, we'd like an expression for the following:

$$\text{Prob}(Y_2^{\alpha_2} | Y_1^{\alpha_1}, F_1^{\alpha_1}, F_2^{\alpha_2}) \quad (25)$$

Using Bayes' Theorem, this becomes

$$\text{Prob}(Y_2^{\alpha_2} | Y_1^{\alpha_1}, F_1^{\alpha_1}, F_2^{\alpha_2}) = \frac{\text{Prob}(Y_1^{\alpha_1}, Y_2^{\alpha_2} | F_1^{\alpha_1}, F_2^{\alpha_2})}{\sum_{Y_2^{\beta_2} \sim F_2^{\alpha_2}} \text{Prob}(Y_1^{\alpha_1}, Y_2^{\beta_2} | F_1^{\alpha_1}, F_2^{\alpha_2})} \quad (26)$$

where $X_2^{\beta_2} \sim F_2^{\alpha_2}$ denotes that the sum is over all possible outcomes corresponding to the measurement $F_2^{\alpha_2}$ (in R_2). (For simplicity, we have suppressed the part of the notation denoting the generalized preparation.) In the causaloid framework, this becomes

$$\text{Prob}(Y_2^{\alpha_2} | Y_1^{\alpha_1}, F_1^{\alpha_1}, F_2^{\alpha_2}) = \frac{\mathbf{r}_{\alpha_1 \alpha_2} \cdot \mathbf{p}}{\mathbf{r}_{\alpha_1 \curvearrowright_2} \cdot \mathbf{p}} \quad (27)$$

where $\mathbf{r}_{\alpha_1 \curvearrowright_2} = \sum_{\beta_2} \mathbf{r}_{\alpha_1 \beta_2}$. (The sum being over β_2 in this notation has the same meaning as the sum being over all outcomes consistent with F_2 .) In order for this probability to be considered well-defined, the right hand side cannot depend on $V - R_1 - R_2$. Since $\mathbf{r}_{\alpha_1 \alpha_2}$ and $\mathbf{r}_{\alpha_1 \curvearrowright_2}$ are determined exclusively by the physical theory, neither has any dependence on $V - R_1 - R_2$. However, \mathbf{p} does depend on $V - R_1 - R_2$. This implies that in order for the probability Eq.(27) to be well defined (i.e. not depend on $V - R_1 - R_2$), it must vary with \mathbf{p} . The dependence on \mathbf{p} can be removed altogether by requiring that $\mathbf{r}_{\alpha_1 \alpha_2}$ be parallel to $\mathbf{r}_{\alpha_1 \curvearrowright_2}$. Therefore, the above probability is well defined if and only if

$$\mathbf{r}_{\alpha_1 \alpha_2} \parallel \mathbf{r}_{\alpha_1 \curvearrowright_2} \quad (28)$$

With this condition, we get

$$\text{Prob}(Y_2^{\alpha_2} | Y_1^{\alpha_1}, F_1^{\alpha_1}, F_2^{\alpha_2}) = \frac{|\mathbf{r}_{\alpha_1 \alpha_2}|}{|\mathbf{r}_{\alpha_1 \curvearrowright_2}|} \quad (29)$$

III. \odot^Γ PRODUCT

Consider two distinct regions; R_A and R_P . By definition

$$\begin{aligned}\mathbf{r}_{\alpha_A \alpha_P} &= \mathbf{r}_{\alpha_A} \otimes^\Lambda \mathbf{r}_{\alpha_P} \\ \mathbf{r}_{\beta_A \alpha_P} &= \mathbf{r}_{\beta_A} \otimes^\Lambda \mathbf{r}_{\alpha_P}\end{aligned}$$

Suppose we wanted to take the dot product between two vectors of the above form. Using decomposition matrices, we can write

$$\begin{aligned}\mathbf{r}_{\alpha_A \alpha_P} \cdot \mathbf{r}_{\beta_A \alpha_P} &= (\mathbf{r}_{\alpha_A} \otimes^\Lambda \mathbf{r}_{\alpha_P}) \cdot (\mathbf{r}_{\beta_A} \otimes^\Lambda \mathbf{r}_{\alpha_P}) \\ &= \sum_{k_A k_P} \left(\sum_{l_A l_P} \Lambda_{l_A l_P}^{k_A k_P} \mathbf{r}_{\alpha_A}|_{l_A} \mathbf{r}_{\alpha_P}|_{l_P} \right) \left(\sum_{l'_A l'_P} \Lambda_{l'_A l'_P}^{k_A k_P} \mathbf{r}_{\beta_A}|_{l'_A} \mathbf{r}_{\alpha_P}|_{l'_P} \right)\end{aligned}\quad (30)$$

where $k_A k_P \in \Omega_{AP}$, $l_A l_P \in \Omega_A \times \Omega_P$, and $l'_A l'_P \in \Omega_A \times \Omega_P$. Notice that we can write

$$\sum_{l_A l_P} \Lambda_{l_A l_P}^{k_A k_P} \mathbf{r}_{\alpha_A}|_{l_A} \mathbf{r}_{\alpha_P}|_{l_P}$$

as

$$\sum_{l_A \in \Omega_A} \left[\mathbf{r}_{\alpha_A}|_{l_A} \left(\sum_{l_P \in \Omega_P} \Lambda_{l_A l_P}^{k_A k_P} \mathbf{r}_{\alpha_P}|_{l_P} \right) \right]$$

Similarly,

$$\sum_{l'_A l'_P} \Lambda_{l'_A l'_P}^{k_A k_P} \mathbf{r}_{\beta_A}|_{l'_A} \mathbf{r}_{\alpha_P}|_{l'_P} = \sum_{l'_A \in \Omega_A} \left[\mathbf{r}_{\beta_A}|_{l'_A} \left(\sum_{l'_P \in \Omega_P} \Lambda_{l'_A l'_P}^{k_A k_P} \mathbf{r}_{\alpha_P}|_{l'_P} \right) \right]$$

Define

$$\Gamma_{l_A}^{k_A k_P}(\mathbf{r}_{\alpha_P}) \equiv \sum_{l_P \in \Omega_P} \Lambda_{l_A l_P}^{k_A k_P} \mathbf{r}_{\alpha_P}|_{l_P}$$

and, similarly,

$$\Gamma_{l'_A}^{k_A k_P}(\mathbf{r}_{\alpha_P}) \equiv \sum_{l'_P \in \Omega_P} \Lambda_{l'_A l'_P}^{k_A k_P} \mathbf{r}_{\alpha_P}|_{l'_P}$$

Using this, Eq.(30) becomes

$$\mathbf{r}_{\alpha_A \alpha_P} \cdot \mathbf{r}_{\beta_A \alpha_P} = \sum_{k_A k_P} \left(\sum_{l_A l'_A} \Gamma_{l_A}^{k_A k_P}(\mathbf{r}_{\alpha_P}) \Gamma_{l'_A}^{k_A k_P}(\mathbf{r}_{\alpha_P}) \mathbf{r}_{\alpha_A}|_{l_A} \mathbf{r}_{\beta_A}|_{l'_A} \right) \quad (31)$$

where $k_A k_P \in \Omega_{AP}$ and $l_A l'_A \in \Omega_A \times \Omega_A$. This suggests that the essence of $\mathbf{r}_{\alpha_A \alpha_P} \cdot \mathbf{r}_{\beta_A \alpha_P}$ is a relationship between \mathbf{r}_{α_A} and \mathbf{r}_{β_A} mediated by matrices that depend on \mathbf{r}_{α_P} . Therefore, we can view Eq.(30) as kind of product of \mathbf{r}_{α_A} and \mathbf{r}_{β_A} . Dot products of this form come up frequently enough that we will define this as the Γ -dot product and denote it as

$$\mathbf{r}_{\alpha_A} \odot^{\Gamma(\mathbf{r}_{\alpha_P})} \mathbf{r}_{\beta_A} \equiv \mathbf{r}_{\alpha_A \alpha_P} \cdot \mathbf{r}_{\beta_A \alpha_P} = (\mathbf{r}_{\alpha_A} \otimes^\Lambda \mathbf{r}_{\alpha_P}) \cdot (\mathbf{r}_{\beta_A} \otimes^\Lambda \mathbf{r}_{\alpha_P}) \quad (32)$$

We will make use of this product later in the paper.

IV. CAUSALLY-UNBIASED ENTROPY

Standard definitions of entropy assume fixed causal structure. Here we develop a causally-unbiased definition of entropy in the causaloid formalism.

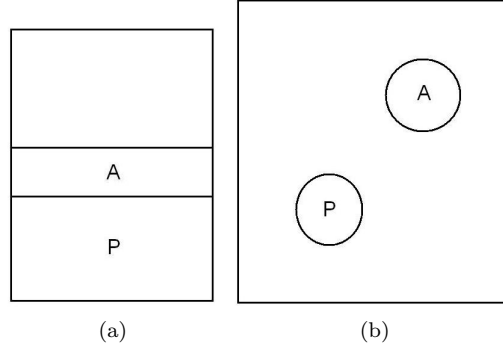


FIG. 2: (a) Fixed causal structure (b) Indefinite causal structure

A. The picture

Shannon entropy [11] for a classical state is defined as

$$S = - \sum_i p_i \log_2 p_i \quad (33)$$

The definition of p_i used in this equation requires that the structure of space-time be organized with the following features:

- a region of interest, A
- an immediate past space-time region, P
- sufficient data about what happened in P
- a measurement F_A
- a set of outcomes, $\{Y_A^i\}$, corresponding to F_A

This allows us to write

$$p_i = \text{Prob}(Y_A^i | F_A, \text{data}_P) \quad (34)$$

Removing all time bias from these features of space-time structure, we get

- a region of interest, A
- a reference region P
- an outcome/measurement pair in P , $\{\alpha_P\} = \{(F_P, Y_P)\}$
- a measurement F_A
- a set of outcomes, $\{Y_A^i\}$, corresponding to F_A

The reference region can be thought of as a kind of preparation region that is not limited to being in the causal past. In fact, the choice of reference region is arbitrary as illustrated in Fig. 2b.

The definition of p_i in a causally-unbiased structure is

$$p_i = \text{Prob}(Y_A^i | Y_P, F_P, F_A) \quad (35)$$

(Since P is arbitrary, we should technically say ' p_i with respect to the reference region P '. However, for the sake of brevity, we will assume that 'with respect to P ' is implied much as 'with respect to the past' is taken as implied in the causally-biased situation.)

Using the above definition of p_i , we define the entropy relative to the reference data (F_P, Y_P) as

$$S = - \sum_i \text{Prob}(Y_A^i | Y_P, F_P, F_A) \log_2 (\text{Prob}(Y_A^i | Y_P, F_P, F_A)) \quad (36)$$

Notice that this reduces to the causally-biased definition of entropy when P is the past; F_A measures the microstate in the classical case or measures in the basis where $\hat{\rho}$ is diagonal in the quantum case.

B. In the causaloid framework

Taking the probability to be well-defined, Eq.(29) and Eq.(36) give the following definition of entropy:

$$S = - \sum_{\alpha_A} \frac{|\mathbf{r}_{\alpha_A \alpha_P}|}{|\mathbf{r}_{\frown_A \alpha_P}|} \log_2 \left(\frac{|\mathbf{r}_{\alpha_A \alpha_P}|}{|\mathbf{r}_{\frown_A \alpha_P}|} \right) \quad (37)$$

Of course, this equation requires that $\mathbf{r}_{\alpha_A \alpha_P} \parallel \mathbf{r}_{\frown_A \alpha_P}$. Loosening this condition slightly, we can consider what happens when $\mathbf{r}_{\alpha_A \alpha_P}$ is nearly parallel to $\mathbf{r}_{\frown_A \alpha_P}$, using the definition of the probability from Eq.(27). The entropy associated with this is

$$S^\Lambda = - \sum_{\alpha_A} \left(\frac{\mathbf{r}_{\alpha_A \alpha_P} \cdot \mathbf{p}}{\mathbf{r}_{\frown_A \alpha_P} \cdot \mathbf{p}} \right) \log_2 \left(\frac{\mathbf{r}_{\alpha_A \alpha_P} \cdot \mathbf{p}}{\mathbf{r}_{\frown_A \alpha_P} \cdot \mathbf{p}} \right) \quad (38)$$

It becomes necessary to shorten the notation for the following work so $\mathbf{r}_{\alpha_A \alpha_P}$ will be denoted as \mathbf{v}_i (where the index α_A is represented by i) and $\mathbf{r}_{\frown_A \alpha_P}$ will be denoted as \mathbf{u} . As with any vector, \mathbf{v}_i can be decomposed into a component parallel to \mathbf{u} and a component perpendicular to \mathbf{u} (i.e. components in $\hat{\mathbf{u}}^\parallel$ and $\hat{\mathbf{u}}^\perp$, respectively). That is,

$$\mathbf{v}_i = v_i^\parallel \hat{\mathbf{u}}^\parallel + v_i^\perp \hat{\mathbf{u}}^\perp \quad (39)$$

Using the unit vectors as defined, \mathbf{p} can be decomposed as

$$\mathbf{p} = p_x \hat{\mathbf{u}}^\parallel + p_y \hat{\mathbf{u}}^\perp + \mathbf{p}^\perp \quad (40)$$

where \mathbf{p}^\perp is the component of \mathbf{p} that is perpendicular to the plane defined by \mathbf{u} and \mathbf{v}_i . The probability of interest, p_i , then becomes

$$\begin{aligned} p_i &= \frac{\mathbf{v}_i \cdot \mathbf{p}}{\mathbf{u} \cdot \mathbf{p}} \\ &= \frac{v_i^\parallel}{u} + k \frac{v_i^\perp}{u} \end{aligned} \quad (41)$$

where $k = \frac{p_y}{p_x}$. Notice that the first term is equivalent to a well-defined probability (Eq. 29). We require the second term to be small since the deviation from well-defined should be small. Since we have already required that \mathbf{v}_i^\perp be small, we need only place restrictions on k .

C. Bounds on k

For the purposes of this subsection, we will work in the plane defined by \mathbf{u} and \mathbf{v}_i . Define the angle between \mathbf{u} and the projection of \mathbf{p} into the plane to be θ . Define the length of the projection of \mathbf{p} into the plane to be p_{xy} . Using basic trigonometry, we get

$$p_y = p_{xy} \sin \theta \quad (42)$$

$$p_x = p_{xy} \cos \theta \quad (43)$$

Therefore, k can be written in a form that is dependent on only one variable, as follows:

$$k = \tan \theta \quad (44)$$

As θ tends towards $\pm \frac{\pi}{2}$, k tends to infinity. Therefore, to ensure that the second term of (41) is small, we require that k be finite. Assume it to be a property of the state space for \mathbf{p} that there exists some $0 < \theta_{max}$. Clearly, $|\theta_{max}| < \frac{\pi}{2}$ in order for k to be finite. So θ is bounded as follows:

$$-\frac{\pi}{2} < -\theta_{max} \leq \theta \leq \theta_{max} < \frac{\pi}{2} \quad (45)$$

The k corresponding to θ_{max} will be denoted as k_{max} . Further bounds can be placed on k by the state space of the physical theory. For our purposes, it is sufficient that k is finite.

D. Q factor

In light of (41), entropy, as defined in (38), becomes

$$\begin{aligned} S^\Lambda &= - \sum_i \left(\frac{v_i^\parallel}{u} + k \frac{v_i^\perp}{u} \right) \log_2 \left(\frac{v_i^\parallel}{u} + k \frac{v_i^\perp}{u} \right) \\ &= - \sum_i \left(\frac{v_i^\parallel}{u} + k \frac{v_i^\perp}{u} \right) \left[\log_2 \left(1 + k \frac{v_i^\perp}{v_i^\parallel} \right) + \log_2 \left(\frac{v_i^\parallel}{u} \right) \right] \end{aligned} \quad (46)$$

Since v_i^\perp is very small (as is implied by the fact that \mathbf{v}_i and \mathbf{u} are nearly parallel) and k is finite, we can take a Taylor expansion (to leading order) of the first \log_2 term. Doing this gives

$$\begin{aligned} S^\Lambda &= - \sum_i \left(\frac{v_i^\parallel}{u} + k \left(\frac{v_i^\perp}{u} \right) \right) \left[\frac{k}{\ln 2} \left(\frac{v_i^\perp}{v_i^\parallel} \right) + \mathcal{O}(v_i^{\perp 2}) + \log_2 \left(\frac{v_i^\parallel}{u} \right) \right] \\ &= - \sum_i \left(\frac{v_i^\parallel}{u} \right) \log_2 \left(\frac{v_i^\parallel}{u} \right) + k \left(\frac{v_i^\perp}{u} \right) \log_2 \left(e \frac{v_i^\parallel}{u} \right) + \mathcal{O}(v_i^{\perp 2}) \end{aligned} \quad (47)$$

Notice that the first term is equivalent to the definition of entropy where $\mathbf{u} \parallel \mathbf{v}_i$ and that S^Λ reduces to this definition when $v_i^\perp = 0$. That is, when $\mathbf{u} \parallel \mathbf{v}_i$ (or equivalently, $v_i^\perp = 0$)

$$S^\Lambda = S \equiv - \sum_i \left(\frac{v_i^\parallel}{u} \right) \log_2 \left(\frac{v_i^\parallel}{u} \right) \quad (48)$$

For $v_i^\perp \neq 0$, we will define

$$Q = - \sum_i \left(\frac{v_i^\perp}{u} \right) \log_2 \left(e \frac{v_i^\parallel}{u} \right) \quad (49)$$

Using k_{max} as defined in the previous section, we can regard $k_{max}Q$ as a kind of correction to the causally-biased entropy. Then, to leading order

$$S - k_{max}Q \leq S^\Lambda \leq S + k_{max}Q \quad (50)$$

E. Understanding Q

Q is an entirely new quantity with no direct classical analogue so understanding its physical interpretation is a non-trivial matter. If we consider entropy as a measure of uncertainty, then S is the measure of our uncertainty that the measurement F_A in region A will yield the specific outcome Y_A^i , given the data we have from the reference region P . Since our reference region P is arbitrary, one way to view Q is that it measures how completely the region P “prepares” region A . In this sense, preparation influences our uncertainty. In a definite causal structure, an immediate past region would completely prepare our region of interest and Q would be zero. However, in the causally-indefinite picture, we cannot require a priori if the reference region that we have chosen will completely prepare our region of interest. If there are no influences on our uncertainty from outside region P , then the probability will be well-defined and Q will be zero. But if there are influences on our uncertainty from outside region P , then the magnitude of Q will reflect that.

F. Using the \odot^Γ product

For the sake of completeness the \mathbf{u} ’s and \mathbf{v}_i ’s must be translated into $\mathbf{r}_{\cap A \alpha_P}$ ’s and $\mathbf{r}_{\alpha_A \alpha_P}$ ’s. Notice that

$$\frac{v_i^\parallel}{u} = \frac{\mathbf{v}_i \cdot \mathbf{u}}{\mathbf{u} \cdot \mathbf{u}} \quad (51)$$

$$\frac{v_i^\perp}{u} = \sqrt{\frac{v_i^2 - v_i^{\parallel 2}}{u^2}} = \sqrt{\frac{\mathbf{v}_i \cdot \mathbf{v}_i - (\mathbf{v}_i \cdot \mathbf{u})^2}{\mathbf{u} \cdot \mathbf{u} - (\mathbf{u} \cdot \mathbf{u})^2}} \quad (52)$$

Substituting $\mathbf{r}_{\alpha_A \alpha_P}$ for \mathbf{v}_i and $\mathbf{r}_{\neg_A \alpha_P}$ for \mathbf{u} gives

$$\frac{v_i^\parallel}{u} = \frac{\mathbf{r}_{\alpha_A \alpha_P} \cdot \mathbf{r}_{\neg_A \alpha_P}}{\mathbf{r}_{\neg_A \alpha_P} \cdot \mathbf{r}_{\neg_A \alpha_P}} \quad (53)$$

$$\frac{v_i^\perp}{u} = \sqrt{\frac{\mathbf{r}_{\alpha_A \alpha_P} \cdot \mathbf{r}_{\alpha_A \alpha_P}}{\mathbf{r}_{\neg_A \alpha_P} \cdot \mathbf{r}_{\neg_A \alpha_P}} - \frac{(\mathbf{r}_{\alpha_A \alpha_P} \cdot \mathbf{r}_{\neg_A \alpha_P})^2}{(\mathbf{r}_{\neg_A \alpha_P} \cdot \mathbf{r}_{\neg_A \alpha_P})^2}} \quad (54)$$

Using the Γ -dot product the above equations simplify to

$$\frac{v_i^\parallel}{u} = \frac{\mathbf{r}_{\alpha_A} \odot^{\Gamma(\mathbf{r}_{\alpha_P})} \mathbf{r}_{\neg_A}}{\mathbf{r}_{\neg_A} \odot^{\Gamma(\mathbf{r}_{\alpha_P})} \mathbf{r}_{\neg_A}} \quad (55)$$

$$\frac{v_i^\perp}{u} = \sqrt{\frac{\mathbf{r}_{\alpha_A} \odot^{\Gamma(\mathbf{r}_{\alpha_P})} \mathbf{r}_{\alpha_A}}{\mathbf{r}_{\neg_A} \odot^{\Gamma(\mathbf{r}_{\alpha_P})} \mathbf{r}_{\neg_A}} - \left(\frac{\mathbf{r}_{\alpha_A} \odot^{\Gamma(\mathbf{r}_{\alpha_P})} \mathbf{r}_{\neg_A}}{\mathbf{r}_{\neg_A} \odot^{\Gamma(\mathbf{r}_{\alpha_P})} \mathbf{r}_{\neg_A}} \right)^2} \quad (56)$$

This allows us to completely specify the entropy of R_A relative to a preparation R_P in the causaloid framework. It is straightforward to generalize this to define the joint entropy of R_A and R_B with reference to a “preparation” R_P . Simply redefine \mathbf{u} and \mathbf{v}_{ij} as

$$\mathbf{v}_{ij} = \mathbf{r}_{\alpha_A \alpha_B \alpha_P} \quad (57)$$

$$\mathbf{u} = \mathbf{r}_{\neg_A \neg_B \alpha_P} \quad (58)$$

where

$$\mathbf{r}_{\alpha_A \alpha_B \alpha_P} = \mathbf{r}_{\alpha_A} \otimes^\Lambda \mathbf{r}_{\alpha_B} \otimes^\Lambda \mathbf{r}_{\alpha_P}$$

and

$$\mathbf{r}_{\neg_A \neg_B \alpha_P} = \sum_{\beta_A} \mathbf{r}_{\beta_A} \otimes^\Lambda \sum_{\beta_B} \mathbf{r}_{\beta_B} \otimes^\Lambda \mathbf{r}_{\alpha_P}$$

Using the same procedure as for one region, we get

$$\frac{v_{ij}^\parallel}{u} = \frac{\mathbf{r}_{\alpha_A \alpha_B} \odot^{\Gamma(\mathbf{r}_{\alpha_P})} \mathbf{r}_{\neg_A \neg_B}}{\mathbf{r}_{\neg_A \neg_B} \odot^{\Gamma(\mathbf{r}_{\alpha_P})} \mathbf{r}_{\neg_A \neg_B}} \quad (59)$$

$$\frac{v_{ij}^\perp}{u} = \sqrt{\frac{\mathbf{r}_{\alpha_A \alpha_B} \odot^{\Gamma(\mathbf{r}_{\alpha_P})} \mathbf{r}_{\alpha_A \alpha_B}}{\mathbf{r}_{\neg_A \neg_B} \odot^{\Gamma(\mathbf{r}_{\alpha_P})} \mathbf{r}_{\neg_A \neg_B}} - \left(\frac{\mathbf{r}_{\alpha_A \alpha_B} \odot^{\Gamma(\mathbf{r}_{\alpha_P})} \mathbf{r}_{\neg_A \neg_B}}{\mathbf{r}_{\neg_A \neg_B} \odot^{\Gamma(\mathbf{r}_{\alpha_P})} \mathbf{r}_{\neg_A \neg_B}} \right)^2} \quad (60)$$

In this manner, we can define causally-unbiased entropy in the causaloid framework for any number of regions.

V. CONCLUSIONS

In a definite causal structure, the only thing required for a definition of entropy that is not in an indefinite causal structure is an immediate past region. Since there is no reason in an indefinite causal structure to choose any reference region over any other, we simply choose an arbitrary region. This ensures that we do not hold on to any pre-conceived notions of space-time and its connection to causality. The definition of the causally-unbiased entropy resulted in a correction to the causally-biased definition of entropy. In a sense, the Q factor gives us an emergent idea of causality. It is a measure of the extent to which our region of interest is causally connected to our reference (or “preparation”) region. If it is zero, the traditional ideas of causality are recovered. The next step would be determining how the Q factor could potentially be physically observed. To do so may require us to know more of the theoretical and mathematical properties of Q. Which mathematical properties of Shannon entropy hold for causally-unbiased entropy? What is the status of the Second Law of Thermodynamics in an indefinite causal structure? To go about answering this, we could consider how S^Λ “evolves” along tubes through indefinite space-times. These questions will be the subjects of continuing work in the near future.

VI. ACKNOWLEDGEMENTS

This work was supported by OGS. Research at Perimeter Institute for Theoretical Physics is supported in part by the Government of Canada through NSERC and by the Province of Ontario through MRI.

-
- [1] R. Penrose and M. A. H. MacCallum, “Twistor theory: An Approach to the quantization of fields and space-time,” *Phys. Rept.* **6**, 241 (1972).
 - [2] S. W. Hawking, “Quantum Gravity And Path Integrals,” *Phys. Rev. D* **18**, 1747 (1978).
 - [3] C. Rovelli and L. Smolin, “Loop Space Representation of Quantum General Relativity,” *Nucl. Phys. B* **331**, 80 (1990).
 - [4] T. Thiemann, “Modern canonical quantum general relativity,” Cambridge, UK: Cambridge Univ. Pr. 819 (2007).
 - [5] R. D. Sorkin, “On the role of time in the sum over histories framework for gravity,” *Int. J. Theor. Phys.* **33**, 523 (1994).
 - [6] J. Ambjorn, J. Jurkiewicz and R. Loll, “Quantum gravity, or the art of building spacetime,” *hep-th/0604212*.
 - [7] L. Hardy, “Probability theories with dynamic causal structure: A new framework for quantum gravity,” *gr-qc/0509120*.
 - [8] L. Hardy, “Towards quantum gravity: A framework for probabilistic theories with non-fixed causal structure,” *J. Phys. A* **40**, 3081 (2007), *gr-qc/0608043*.
 - [9] L. Hardy, “Formalism Locality in Quantum Theory and Quantum Gravity,” *gr-qc/0804.0054*.
 - [10] L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics Vol. 5: Statistical Physics Pt. 1* 3rd Ed. (Butterworth Heinemann, Oxford, 1980).
 - [11] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, Cambridge, 2000).